Higher Education Digitisation Service: access in the future, preserving the past - the UK perspective.

Simon Tanner - Digitisation Consultant, HEDS, University of Hertfordshire.

Brian Robinson - Manager, HEDS, University of Hertfordshire.

Paper for the Retrospektive Digitalisierung von Bibliotheksbestanden

Workshop, 26-27 January 1998, Gottingen, Germany.

Summary

This paper will introduce HEDS and our services, with focus on how HEDS fits into the overall digital library initiatives of UK higher education.

Introduction

The aim of the Higher Education Digitisation Service (HEDS) is to establish a range of core and value-added services available through a single point of contact to support the conversion of high volumes of learning, teaching and scholarly materials into electronic forms for increased availability in the higher education community. The University of Hertfordshire has established HEDS by bringing together a wide range of expertise and specialisms with project funding from the Electronic Libraries Programme (eLib). This project funding was initially granted for 2 years from September 1996.

HEDS was established and funded because of a definite requirement within the UK higher education community for digitisation services. The driving force behind the development of such services came out of the recommendations of the Higher Education Joint Funding Councils' Libraries Review Group (Follett Report) published in December 1993 for the exploitation of IT in future higher education library provision. The conversion of printed texts into electronic form was seen to have potential particularly in three areas: space-saving, preservation and for improved access. As a result of the Follett Report's range of recommendations for exploitation of IT, the national Electronic Libraries Programme (eLib) with a budget of £15 million was set up under the Joint Information Systems Committee (JISC). It comprises many projects in key areas such as document delivery, on-demand publishing, electronic journals and access to networked information resources. Some projects aim to investigate the feasibility of alternative ways of working and providing information. Others, like HEDS, have initiation funding for what is intended to become an established long term service. Many of the Electronic Libraries Programme projects have experienced difficulties, not in creating digital library resources, but in populating those resources with large enough amounts of source information. Larger scale conversion infrastructure could only be effectively achieved by a centrally resourced national service - HEDS. HEDS has already worked with the majority of the current eLib projects either providing technical advice or offering conversion services.

The HEDS project plan has three phases: to establish and then offer digitisation services, followed by phased transition to a self supporting service. In Year 1, from September 1996, HEDS was established, equipped and staffed. A number of digitisation projects, also funded by eLib, are progressing well to 'pipeclean' the HEDS processes and services. HEDS has been able to establish effective business processes and test our technical capabilities through these projects. In Year 2, HEDS is now open for business and ready to accept digitisation work by request from a range of

clients, both from higher education institutions and from other sources. In Year 3, HEDS will begin the transition to a self supporting service through the implementation of a business plan agreed with the JISC.

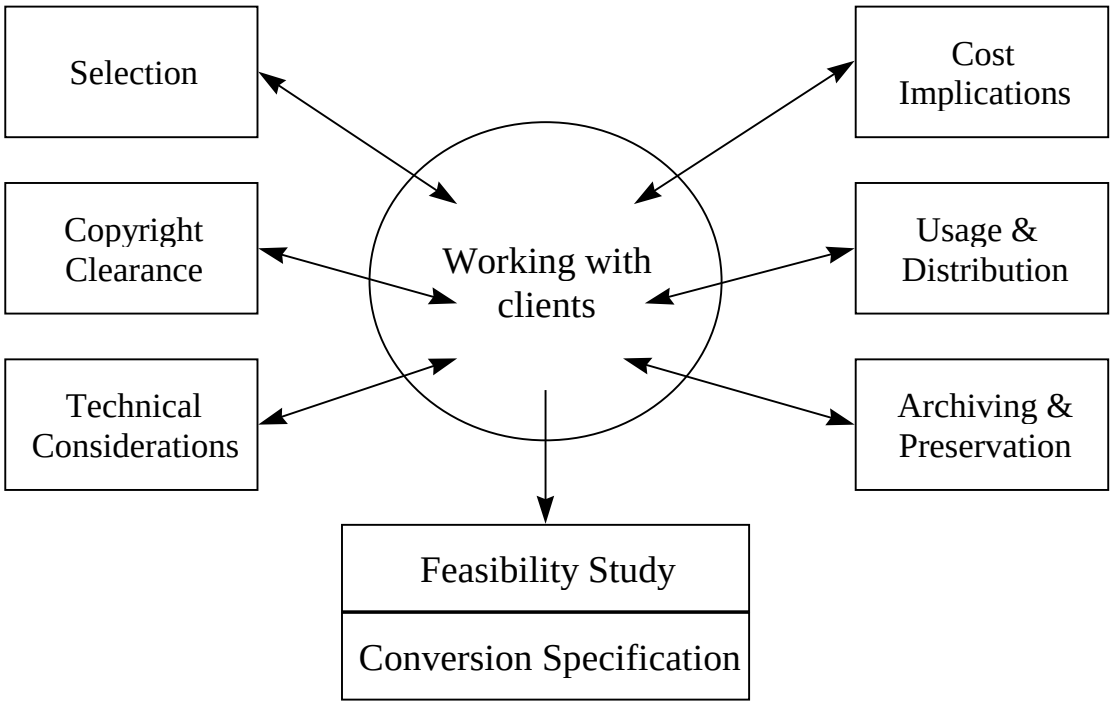
Services

The main objective in the HEDS service plan is to provide a total management package so that our clients have a single point of access to a range of co-ordinated and interlocking services. HEDS will work with clients to agree and deliver a complete package, including all the issues identified in Figure 1 below.

At this stage in the service development, HEDS is focusing on the following core services.

- Advice and consultancy to clients on the feasibility of digitising defined collections of materials.
- Guidance on selecting the most cost-effective methods for realising your digitisation aims.
- To manage the complete job, from problem definition to final product delivery and acceptance.
- To prepare functional and technical conversion specifications for the digitisation work.
- To provide quality assurance procedures to validate the end product.
- To deliver digitised materials on time and to agreed standards.
- HEDS will also provide advice and assistance with copyright clearance.
- HEDS will take a central role in raising awareness of digitisation for the higher education sector.

Figure 1: HEDS Service Development



HEDS sees its role as being able to advise and carry out production to support all of the elements identified in Figure 1 and especially to extend the specialist services. HEDS has drawn together extensive experience in handling valuable, unique and fragile materials with appropriate care and security. One of the obvious benefits of creating electronic versions of such materials is the preservation through alternative access reducing wear and tear on the original. HEDS also provide copyright research in connection with some projects. Further development will support the provision of copyright clearance services and referral to expert legal advice, where appropriate, as value-added services.

Conversion Projects

HEDS has a number of eLib funded projects in progress at the moment which will result in much wider access to the materials with more than 105,000 pages converted. These projects include the following materials:

- Council of British Archaeology research reports.
- Transactions of the Institute of British Geographers.
- Meteorological observatory data (1881 - 1975).
- The Statistical Accounts for Scotland (1799 & 1845).
- Social policy and transport pamphlet collection of the British Library of Political and Economic Science.

These will present a valuable resource to the higher education community and a number of technical challenges for HEDS in converting a range of materials into various electronic formats. The original materials include 35mm microfilm, paper sizes from A5 to greater than A3 with handwriting, printed text and graphics including photographs in journal and pamphlet formats, bound volumes and single sheets. The electronic output has required techniques including image scanning, optical character recognition, conversion to Adobe Acrobat PDF and rekeying of some indexes.

Consultancy Projects

HEDS has also been working on feasibility studies with a number of Universities to establish the techniques and costs of digitisation for some large projects. It is hoped that many of these will lead to full digitisation projects later this year.

- World's Fair Newspaper - University of Sheffield.
  A project to discover the costs and appropriate techniques for conversion and indexing of approximately 350,000 pages of this weekly newspaper from 1904 onwards. Samples from microfilm and broad sheet originals were made and assessed.
- Refugee Studies Programme - Digital Library project - Oxford University.
  A feasibility study to deliver a specification for the technical requirements of converting this unique collection of internationally important grey literature and reports into appropriate digital formats. This involves extensive sampling and also a vertical slice survey of approximately 1,000 items through the collection to ascertain the exact proportions of the collections resources.

- JIDI (JISC Image Digitisation Initiative).
  A feasibility study to deliver specifications of the technical requirements, benchmarks and costs for twenty projects which make up the arts and humanities oriented collection to be converted. These projects include a diverse range of resource formats including; 35mm colour slides, photographic plates, paintings, posters, protest banners and geological samples among the more challenging materials. The solutions require a single master image with associated metadata of both a technical and subject descriptive nature.

- Pilot of Nature.
  This pilot study is to establish the techniques and standards for the possible digitisation of the journal Nature (1869 - 1992). This pilot is funded by JISC's Committee on Electronic Information. Macmillan Publishers Ltd the owners of the magazine will make the source material available. Initially, there will be several hundred pages converted and made available to UK Higher Education as part of an evaluation. The University of Manchester Computing Centre will be mounting the data. HEDS will determine the costs of digitisation to an acceptable standard and meet the challenges that are likely to arise when digitising the different fonts, styles and paper that Nature has utilised over the past 130 years. This pilot is in the very early stages, but is expected to follow the general model established by JSTOR of images for each page viewed on screen with hidden searchable text. Metadata will be created down to each logical item (such as an article or letter) and the pilot partners are experimenting with embedding this metadata within the image files.

Expanding the Technology Envelope

HEDS is a service provider which is attempting to use all the latest technologies and techniques to enable it to deal with a wide variety of formats at higher standards. However, the state of the technology available is such that much of our added value work is unacceptably constrained by the sheer cost of creating the images from originals in difficult formats. In this category I would include bound volumes to archive standards, some microfilm and colour photographic resources. The issue for HEDS is to get the balance between the cost of conversion, the available technology and our application of that technology correct. Previously, costs have constrained medium volume, high standard scanning because of the amount of human intervention required to get acceptable images from the available technology. New technology and techniques are making this intervention smaller and thus will assist in driving costs down and standards up.

One of the best examples of this trend has been scanning from bound volumes to archival standards. In the UK there is quite a demand for high standards of scanning from 18th and 19th Century bound volumes. These must be completed with the absolute minimum of damage to the originals which are often unique and always fragile. Thus, flat bed scanning is not acceptable and an overhead scanner with a book cradle is required. HEDS purchased a high specification bookscanner, the Zeutschel Omniscan 5000, for the express purpose of being able to cope with this type of resource. Certainly this equipment is very advanced for this type of medium and being able to take scans of originals up to A1 size is exceptional. However, this standard of equipment is still very young and the number of manufacturers of such equipment may be counted on one hand. This youth is indicated by the challenge that true 256 greyscale scanning has been to manufacturers of bookscanners, with Zeutschel only being able to present 400 dpi greyscale capabilities in the last quarter of 1997. This is exactly the sort of capability that is now in demand in the UK Higher Education community but has not previously been supported by suitable equipment. The next challenge for bookscanner manufacturers is to get the resolutions higher for larger document sizes. These sorts of developments are essential to enable the costs of scanning to be driven down. Scanning must be made easier through removing handling difficulties and thus reducing damage to a minimum, whilst also reducing operator intervention to gain good images from the equipment and software available.

A similar problem has been faced when approaching microfilm resources. These are ideal when the original is either no longer available, is in too fragile or poor condition for scanning or where the size of the original is too big for available equipment. However, there are again only a very few manufacturers of microfilm scanners capable of the highest standards of scanning greyscale images at 400 dpi. As far as I am aware there are no microfilm scanners capable of production at 600 dpi.

HEDS is quite aware that the higher resolutions I have expressed a need for are not required in all cases. Certainly, the number of times that HEDS is asked to produce 600 dpi images is much higher than the amount of times HEDS actually recommends this option. Frequently, HEDS are being asked to offer JSTOR equivalent solutions to UK Higher Education. JSTOR have a strategy of scanning all their materials at 600 dpi, but have the advantage of always being able to strip the originals for flatbed scanning. However, our clients do not often have either the financial resources or the technical need to convert items at such a high resolution. HEDS also finds, in most of our projects, that stripping the originals is not a desired option. Again, HEDS are trying to get the balance between cost, image standards and file sizes balanced to provide an optimum solution.

I have spoken about the technology lag in some areas of digitisation. But there is one area where the technology is well established, but our capability to apply the technology has not been perfected. The area I refer to is colour originals and in particular photographic materials. The main difficulties are not producing good images, but in the Quality Assurance and benchmarking of these images. Producing a single standard for a collection of photographic materials is a real challenge as each item has more individual variety than, for instance, a book. The photographs subject, the lighting, the camera exposure, the colour filters used, the film type, the reproduction of the negative, the standard of the photographic print or 35mm slide - these are all variables in every single colour photographic original. Similarly in the scanning process there are a large number of variables, the colour registration, the operators screen standard and room lighting, the equipment and software used and the image output standard specified. All these factors conspire to make the process more subjective and less quantifiable. Therefore, benchmark sample images, colourbars and Quality Assurance become even more important. HEDS has followed much of the advice of the University of Virginia's Electronic Text Center (http://www.village.virginia.edu/iath/treport/scanfilm.html) to develop our own standards for benchmarking in this area. We will be testing these in the JIDI project and will report our findings on our Web site (http://heds.herts.ac.uk).

The Future of Digitisation in the UK

There are a number of issues that I think will be important to ensure the continued development of digital library resources in the UK.

- **Continued investment in the conversion infrastructure.**
  Resources are scarce and funds will be best spent on developing centralised resources or using vendors, rather than every project purchasing equipment with limited lifetimes and funds.

- **Technology Catch Up.**
  The equipment and software needed to achieve the high standards of conversion required to serve the Higher Education communities needs will become available and ever more efficient with less and less human intervention required. This will drive down the cost of conversion whilst increasing the standards of images. This will encourage further projects to digitise materials previously inaccessible to conversion.

- **Maximise value of materials.**
  There is a trend towards high profile, large scale conversion projects that will benefit all of Higher Education in the UK. These are extremely important, but I hope that the funding bodies do not remove funding from the smaller projects with a very focused user group in mind where the costs are low and the value added very high.

- **Planned growth of digital libraries.**
  The Electronic Libraries Programme under JISC in the UK has developed many resources to create digital libraries. However, much of this has come about through the strong interests and work of a few people without institutional review. There is a need to continue to develop the top level planning to resolve digitisation priorities based on experts within disciplines determining the core resources that would be most valuable for teaching.

- **Scalability.**
  Again the Electronic Libraries Programme has been successful in creating small scale digital libraries. However, as the digital resources grow we must plan to ensure that there is sufficient capacity to manage and maintain large sites and large electronic collections.

- **Access Issues and Metadata.**
  The future costs for conversion will continue to drop, but the costs of creating intellectual access and metadata will remain high. It is important to develop appropriate levels of content description to enable access mechanisms if the value of the digital library resources are to be fully realised.

HEDS: Futures and Key Benefits

In a short time HEDS has built up a wide ranging service and gained experience which will enable it to meet future clients' requirements. The main barrier facing most digitisation projects is cost. Digitisation has for too long been associated either with high costs or low-grade end products. HEDS has the benefit of its project funding to ensure services are offered to HEIs at well below commercial rates. HEDS is also striving to provide a high-grade service that matches the detail and standards of the end product to the vision of the client's requirements.

In conclusion, HEDS believes that the key benefits that its services offer are:

- The single point of contact for access to a range of digitisation services and management of the total package.

- The flexibility to tailor conversion advice, specification and delivery to suit the job in hand.

- Competitive rates including heavily subsidised rates for HEIs during the project period.

- Expert advice, consultancy and project management backed up by timely, reliable, quality conversion.